Fig. 4 is a flow chart of a process in a pattern by-characteristic pattern selection unit in coincidence with the first exemplary embodiment.

Fig. 5 is a flow chart of a process in a speaker adaptation processor in coincidence with the first exemplary embodiment.

Fig. 6 is a block diagram of a voice controller in coincidence with a second exemplary embodiment of the present invention.

Fig. 7 is a block diagram of a voice controller in coincidence with a third exemplary embodiment of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Embodiments of the present invention will be described hereinafter with reference to the accompanying drawings.

Exemplary embodiment 1

A first exemplary embodiment is shown in Fig. 1 through Fig. 5. A user speaks a word defined for pattern selection, speaker adaptation, and device selection (hereinafter called a pattern selection word). Voice controller 102 receives a speech sound of the word through microphone 101. The user subsequently speaks a word for indicating a content of controlling the device (hereinafter called a device control word) selected by the pattern selection word. When voice controller 102 receives a speech sound of the word, it outputs a control signal to the selected device.

In the present embodiment, for example, the user speaks "Television_Increase-sound" or "Light_Turn off". The utterance includes "Television" or "Light" as a pattern selection word, and "Increase sound" or "Turn off" as a device control word. Based on the utterance, voice controller 102 sends to television 103 a control signal for increasing its sound volume, or to lighting 104 a control signal for turn-off.

An operation in the voice controller shown in Fig. 2 will be described hereinafter in detail.

A speech signal of a pattern selection word, namely a first utterance, fed through the microphone is converted from an analog signal to a digital signal by sound input unit 201, and supplied to acoustic analysis unit 202. Acoustic analysis unit 202 determines a linear predictive coding (LPC) cepstral coefficient vector as acoustic parameters of the speech sound digital signal. The present embodiment provides the example using the LPC cepstral coefficient vector as the acoustic parameters, but the other acoustic

parameters such as mel frequency cepstral coefficients (MFCC) produce a similar advantage.

Fig. 3 shows a detailed configuration of pattern by-characteristic storage 203.

The learned speech sound data stored in pattern storage 203 is previously categorized according to ages of speakers. Patterns by characteristic previously stored in pattern by-characteristic storage 203 comprise the following data:

average values of LPC cepstral coefficient vectors of utterances classified by every phonetical unit of each time by training speakers in each characteristic category;

covariance values of LPC cepstral coefficient vectors of every utterance by training speakers in each characteristic category;

average values of LPC cepstral coefficient vectors of utterances classified by every phonemes' state of each time by all training speakers; and

covariance values of LPC cepstral coefficient vectors of every utterance by all training speakers.

The present embodiment uses the following three categories:

	Ages
Category 1	– 12
Category 2	13 – 64
Category 3	65 –

Pattern storage 203 stores trained patterns including the following data;

average values 301 of every utterance by all training speakers;

covariance values 302 of every utterance by all training speakers;

average values 311 of utterances by training speakers in category 1;

average values 312 of utterances by training speakers in category 2;

average values 313 of utterances by training speakers in category 3;

covariance values 321 of the utterances by the training speakers in category 1;

covariance values 322 of the utterances by the training speakers in category 2; and

covariance values 323 of the utterances by the training speakers in category 3.

The LPC cepstral coefficient vector is determined by acoustic analysis unit 202. Pattern by-characteristic selection unit 204 performs a distance

calculation of a first utterance part of the LPC cepstral coefficient vector using previously prepared trained patterns in pattern storage 203. Based on the resultant calculation, selection unit 204 further determines a pattern by characteristic to be used for recognizing the subsequent word. This calculation uses a linear function developed from Mahalanobis' distance as a distance measure (similarity). Mahalanobis' distance is a fundamental function, and the developed linear function is called a simplified Mahalanobis' distance. The simplified Mahalanobis' distance is disclosed in U.S Patent Number 4,991,216. The present embodiment uses a statistical distance measure, namely the simplified Mahalanobis' distance; however, another statistical distance measure such as Bayes' discriminant may be used. An output probability of hidden Markov model may also be used to produce the similar advantage.

Fig. 4 is a flow chart of a process in pattern selection unit 204. Acoustic parameters of an input speech sound is read in the first step (step S401). In other words, an LPC cepstral coefficient vector obtained by acoustic analysis of the input speech sound is read in the present embodiment.

Next, patterns to be selected are read (step S402). In the present embodiment, the patterns are read from patterns stored on pattern storage 203 shown in Fig. 3. Average values 311 of utterances by training speakers in category 1, average values 301 of all utterances by all training speakers, and covariance values 302 of all utterances by all training speakers are read for category 1. Average values 312 of utterances by training speakers in category 2, average values 301 of all utterances by all training speakers, and covariance values 302 of all utterances by all training speakers are read for category 2. Average values 313 of utterances by training speakers in category 3, average values 301 of all utterances by all training speakers, and covariance values 302 of all utterances by all training speakers are read for category 3.

A distance calculation of a pattern selection word is then performed using each trained pattern (step S403). The word is used both as a selection of trained patterns to be used for recognizing the following word and a selection of a device. The distance calculation determines a distance between defined all pattern selection words and the input speech sound by a user for each trained pattern for each category read in step S402.

The distance calculation uses equation (1) of the simplified